A strategic level for scientific digital libraries

Feng, L.; Jeusfeld, M.A.; Hoppenbrouwers, J.J.A.C.

Published in:
Digital libraries: People, knowledge, and technology: Proceedings of the 5th international conference on Asian digital libraries, ICADL 2002, Singapore, December 11-14, 2002

Publication date:
2002

Citation for published version (APA):
Feng, L., Jeusfeld, M. A., & Hoppenbrouwers, J. J. A. C. (2002). A strategic level for scientific digital libraries. In E-P. Lim (Ed.), *Digital libraries: People, knowledge, and technology: Proceedings of the 5th international conference on Asian digital libraries, ICADL 2002, Singapore, December 11-14, 2002* (Vol. 5, pp. 495-506). (Lecture notes in computer science; Vol. 5, No. 2555). Springer Verlag.

A Strategic Level for Scientific Digital Libraries

Ling Feng[1] , Manfred Jeusfeld[2], and Jeroen Hoppenbrouwers[2]

[1] Dept. of Computer Science, University of Twente, PO Box 217
7500 AE Enschede, The Netherlands. ling@cs.utwente.nl[1]
[2] Infolab, Tilburg University, PO Box 90153
5000 LE Tilburg, The Netherlands. {jeusfeld,hoppie}@kub.nl

Abstract. Digital libraries (DLs) are a resource for answering complex questions. Up to now, such systems mainly support keyword-based searching and browsing. The mapping from a research question to keywords and the assessment whether an article is relevant for a research question is completely with the user. In this paper, we present a two-layered DL model. The aim is to enhance current DLs to support different levels of human cognitive acts, thus enabling new kinds of knowledge exchange among library users. The low layer of the model, namely, the *tactical cognition support layer*, intends to provide users with requested relevant documents, as what searching and browsing do. The upper layer of the model, namely, the *strategic cognition support layer*, not only provides users with relevant documents but also directly and intelligently answers users' cognitive questions. On the basis of the proposed model, we divide the DL information space into two subspaces, i.e., a *knowledge subspace* and a *document subspace*, where documents in the document subspace serves as the justification for the corresponding knowledge in the knowledge subspace. Detailed description of the knowledge subspace is particularly discussed.

1. Introduction

Users' information retrieval activities have traditionally been categorized into *searching* and *browsing*. Searching implies that users know exactly what to look for, while browsing should assist users navigating among correlated searchable terms to look for something new or interesting. So far, most of the major work on DLs focuses on supporting these two kinds of information requirements. To support efficient searching activities, efforts have been made in developing retrieval models, building document and index spaces, extending and refining queries for DLs [9]. In [6], index terms are automatically extracted from documents and a vector - space paradigm is exploited to measure the matching degrees between queries and documents. Indexes and metadata can also be manually created from which semantic relationships are captured [2]. Furthermore, the information space consisting of a large collection of documents can be semantically partitioned into different clusters, so that queries can be evaluated against relevant clusters [21]. According to topic areas, a distributed

[1] This work was completed while the author was in the Infolab of Tilburg University in the Netherlands.

semantic framework is proposed in [17] to contextualize the entire collection of documents for efficient large-scale searching. To improve query recall and precision, several query expansion and refinement techniques based on relational lexicons/thesauri or relevance feedback have been explored [20].

Since one DL usually contains lots of distributed and heterogeneous repositories which may be autonomously managed by different organizations, in order to facilitate users' browsing activities across diverse sources easily, many efforts have been engaged in handling various structural and semantics variations and providing users with a coherent view of a massive amount of information. Nowadays, the interoperability problem has sparked vigorous discussions in the DL community. The concept extraction, mapping and switching techniques enable users in a certain area to easily search the specialized terminology of another area. A dynamic mediator infrastructure [13] allows mediators to be composed from a set of modules, each implementing a particular mediation function, such as protocol translation, query translation, or result merging [15]. [18] presents an extensible digital object and repository architecture FEDORA, which can support the aggregation of mixed distributed data into complex objects, and associate multiple content disseminations with these objects. [16] employs the distributed object technology to cope with interoperability among heterogeneous resources. With XML becoming the Web data exchange standard, considerable work on modeling, querying and managing semistructured data and non-standard data formats are conducted to enable the integration of heterogeneous resources [3,7].

Despite lots of fruitful achievements in the DL area, from the standpoint of satisfying human's information needs, the current DL systems suffer from the following two shortcomings.

Inadequate High-Level Cognition Support. The traditional use of DLs is keyword-based. Users request information by entering some keywords, and DL systems return matching documents. But users expect more than this. Typically, users have some pre-conceived hypotheses or domain-specific knowledge. They may desire the library to confirm/deny their existing hypotheses, or to check whether there are some exceptional/contradictory documental evidences against the pre-existing notions, or to provide some predictive information so that they can take effective actions. For example, a user working in a flood-precaution office is concerned about whether there will be floods in the coming summer. According to his previous experience, it seems that "*A wet winter may cause floods in summer*". In this situation, instead of using *disperse keywords* to ask for *documents*, the user would prefer to pose a *direct question* to DLs like "*Does a wet winter cause floods in summer*?" and expect a confirmed/denied *intelligent answer* as well as a series of *supporting literatures* to justify the answer, rather than a list of articles lacking explanatory semantics and waiting for his further checking.

Inadequate Knowledge Sharing and Exchange Channel. Traditional libraries are a public place where a large extent of mutual learning, knowledge sharing and exchange can happen. A user may ask a librarian for searching assistance. Librarians may collaborate in the process of managing, organizing and disseminating information. Users themselves may communicate and help each other in using library resources. When we progress from physical libraries to virtual DLs, these valuable features must be retained. Future DLs should not just be simple storage and archival systems. To be successful, DLs must become a *knowledge place* for a wide spread of knowledge acquisition, sharing and propagation. In the above example, if the DL could make readily available knowledge and expertise to the public users, which might otherwise require time-consuming searching and consultation with librarians and/or experts, we can improve users' working effectiveness and efficiency. Also, as machine knowledge does not deteriorate with time as that human knowledge does, for long-term retention, DLs provide an ideal repository for the knowledge in the world.

In this paper, we propose a two-layered DL model to support users' tactical and strategic level information requirements. The model moves beyond simple searching and browsing across multiple correlated repositories, to acquisition of knowledge. On the basis of the proposed function model, we further divide the DL information space into two subspaces, i.e., a *knowledge subspace* and a *document subspace*. Documents in the document subspace serves as the justification for the corresponding knowledge in the knowledge subspace.

The remainder of the paper is organized as follows. In section 2, we outline a two-layered DL model. A formal description of the DL knowledge subspace is presented in section 3. Section 4 concludes the paper.

2. A Two-Layered DL Function Model

We propose a two-layered DL model, consisting of a *tactical cognition support layer* and a *strategic cognition support layer*, to address users' information needs, as shown in Figure 1.

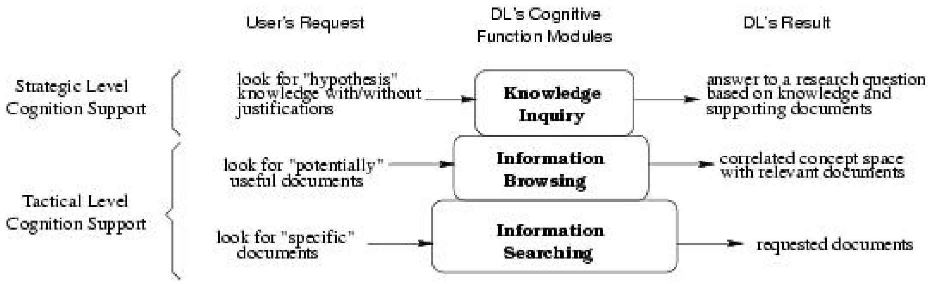

Fig. 1. A two-layered DL model

2.1 Tactical Level vs. Strategic Level Cognition Support

Tactical level cognition support. We view traditional DL searching and browsing as tactical level cognitive acts. The target of searching is towards certain specific documents. One searching example is "*Look for the article written by John Brown in the proceedings of VLDB88*." As the user's request can be precisely stated beforehand, identifying the target repository where the requested document is located is relatively easy. In comparison to searching whose objective is well-defined, browsing aims to provide users with a conceptual map, so that users can navigate among correlated items to hopefully find some potentially useful documents, or to formulate a more precise retrieval request further. For instance, *a user reads an article talking about a water reservoir construction plan in a certain region. S/he wants to know the possible influence on ecological balance. By following semantic links for the water reservoir plan in the DL, s/he navigates to the related ‘‘ecological protection" theme, under which a set of searchable terms with relevant documents are listed for selection*. As the techniques of searching and browsing have been extensively studied and published in the literature, we will not discuss these any further here.

Strategic level cognition support. In contrast to tactical level cognition support which intends to provide users with requested documents, strategic level cognition support not only provides relevant documents but also intelligently answers high-order cognitive questions, and meanwhile provides justifications and evidences. For instance, instead of retrieving documents with dispersed keywords like *wet winter* and *summer flood*, etc, the user would prefer to pose a direct question like "*Tell me whether a wet winter will cause summer flood*", and expect a direct confirmed/denied answer from the DL system rather than a list of articles lacking explanatory semantics and waiting for his assessment. The provision of strategic level cognition support adds values to DLs beyond simply providing document access. It reinforces the exploration and utilization of information in DLs, and advocates a more close and powerful interaction between users and DL systems.

2.2 An Enlarged DL Information Space

In order to support the two kinds of cognitive acts, we further divide the DL information space into two subspaces, i.e., a *knowledge subspace* and a *document subspace*, as shown in Figure 2. Here, we only illustrate the organization of documents for knowledge justification purpose. Documents in a DL are in fact also indexed and clustered based on the ontology and thesauri.

The knowledge subspace. The basic constituent of the *knowledge subspace* is knowledge, such as hypothesis, rule, belief, etc. In this initial study, we focus on hypothesis knowledge in empirical sciences. Each hypothesis describes a certain relationship among a set of concepts. For example, the hypothesis "H_2: *wet winter causes summer flood*" explicates a causal relationship between a cause *wet winter* and the effect *summer flood* it has. Considering that the DL knowledge subspace is for users to retrieve strategic level knowledge, it will inevitably be subject to the classical information retrieval's vocabulary (synonymy and polysemy) problem. Previous research [1,5] demonstrated that different users tend to use different terms to seek identical information. To enable knowledge exchange and reusability, we build various relationships including *equivalence, specification/generalization* and *opposition* over the hypotheses knowledge, expanding a single user's hypothesis into a network of related hypotheses. Later, if one user's inquiry has the form of a hypothesis, the above relationships can be explored to find matching hypotheses in the knowledge subspace. The hypotheses together with the backing documents, serving as the justification of hypotheses, in the document subspace are returned to the user as a part of the answer to his/her strategic request. For example, a more general hypothesis in respect to H_2 is "H_1: *wet winter is related with river behavior*".

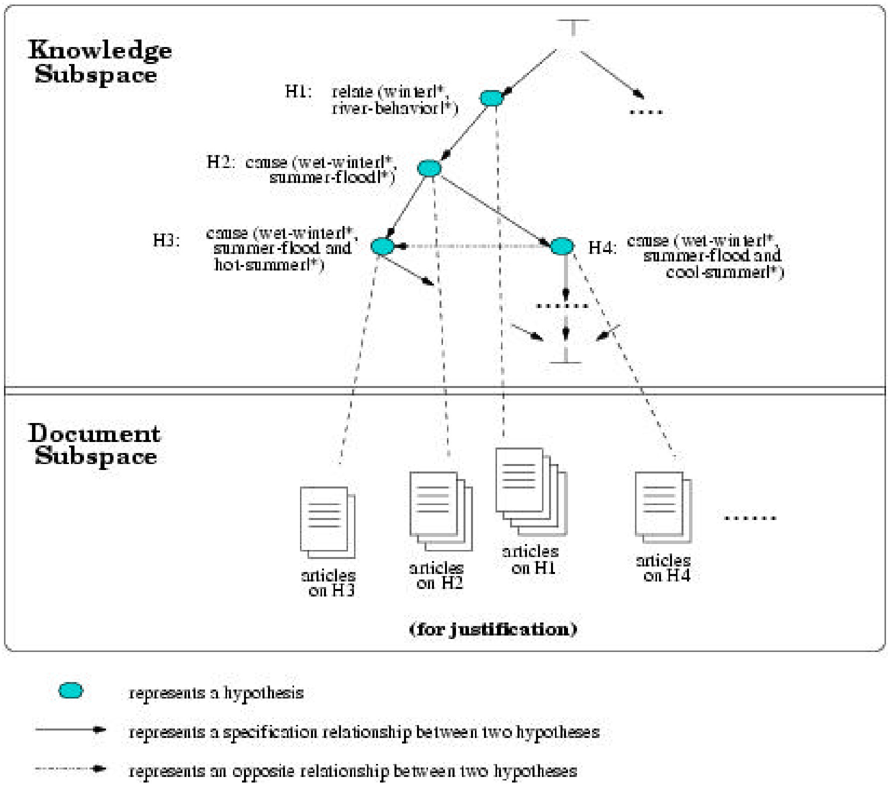

Fig. 2. An enlarged DL information space

The document subspace. Under each hypothesis is a justification set, giving reasons and evidences for the knowledge. These justifications, made up of articles, reports, etc., constitute the document subspace of the DL information space. Taking the above hypothesis H2 for example, the articles mentioning exactly that "*wet winter is an indicator of summer flood*." constitute the justification for that hypothesis. It is worth notice here that the document subspace challenges traditional DLs on literature organization, classification, and management. For belief justifications, we must extend the classical *keyword*-based index schema, which is mainly used for information searching and browsing purposes, to *knowledge*-based index schema, in order that the information in DLs can be easily retrieved by both keywords and knowledge.

3. A Formal Description of DL Knowledge Subspace

In this section, we define the basic constituent of the DL knowledge subspace - *hypothesis*, starting with its two constructional elements, i.e., *concept terms* and *relation terms*. Throughout the discussion, we assume the following notation is used.

- A finite set of entity concepts $EConcept=\{e_1, e_2, \ldots, e_t\}$.
- A finite set of relation concepts $RConcept=\{r_1, r_2, \ldots, r_s\}$.
- A finite set of concepts *Concept*, where $Concept = EConcept \cup RConcept$.
- A finite set of contextual attributes $Att=\{a_1, a_2, \ldots, a_m\}$. The domain of $a_i \in Att$ is denoted as $Dom(a_i)$.
- A finite set of concept terms $CTerm = \{n_1, n_2, \ldots, n_u\}$.
- A finite set of relation terms $RTerm = \{m_1, m_2, \ldots, m_v\}$.
- A finite set of hypotheses $Hypo = \{H_1, H_2, \ldots, H_w\}$.

3.1 Concepts

Concepts represent real-world entities, relations, states and events. We classify concepts into two categories, i.e., *entity concepts* and *relation concepts*. Entity concepts are used to describe concept terms, while relation concepts are mainly for presenting various relationships among concept terms. Based on the substantial work on lexicography and ontology [19,12,8,14,10,11], three typical primitive relationships (*Is-A*, *Synonym*, and *Opposite*) between concepts can be established.

3.2 Concept Terms

Each entity concept can be associated with a conceptual context, denoting the circumstance under which the entity concept is considered. Basically, a conceptual context can be described using a set of attributes called **contextual attributes**, each of which represents a dimension in the real world. Typical contextual attributes include *time*, *space*, *temperature*, and so on. For example, the context for the entity concepts *wet-winter* and *summer-flood* could be constructed by two contextual attributes, *Year* and *Region*. Such a context can be further instantiated by assigning concrete values to its constructional attributes. For example, we can restrict the contextual *Region* and *Year* of *wet-winter* to the *north* and *south* areas in *2000* by setting *Region:={"north", "south"}* and *Year:={"2000"}*. *Region:=Dom*(*Region*) assigns all applicable regions ("*south*", "*north*", "*east*", and "*west*") to contextual attribute *Region*. In this paper, we name an entity concept with an associated context as a **concept term**.

Definition 1. A **concept term** n is of the form $n = e|_{AV}$, where $e \in EConcept$ and $AV= \{a := V_a \mid (a \in Att) \wedge (V_a \subseteq Dom(a))\}$.
Let $Att = \{a_1, a_2, \ldots, a_m\}$ be a set of contextual attributes which constitute the context under consideration. The default setting for attribute $a_i \in Att$ is the whole set of applicable values in the domain of a_i, i.e., a_i:=$Dom(a_i)$. $AV=\{a_1$:=$Dom(a_1)$,

a_2:=$Dom(a_2)$, …, a_m:=$Dom(a_m)$} depicts a universe context. A simple and equivalent representation of the universe context is $AV = *$.

Example 1. Suppose the context is comprised of two contextual attributes *Region* and *Year*. *wet-winter*$|_{\{Region:=\{“north”,\ “south”\},Year:=\{“2000”\}\}}$ is a concept term, denoting a *wet-winter* entity concept in the *north* or *south* in *2000*.
wet-winter$|_{\{Region:=Dom(Region),Year:=Dom(Year)\}}$ is equivalent to *wet-winter*$|_*$ according to Definition 1.

Different relationships of concept terms can be identified based on their entity concepts and associated contexts. Before giving the formal definitions, we first define three relationships between two instantiated contexts.

Definition 2. Let AV_1 and AV_2 be two instantiated contexts.

- $AV_1 \leq_a AV_2$ (or $AV_2 \geq_a AV_1$), iff $\forall(a:= V_2) \in AV_2\ \exists(a:= V_1) \in AV_1\ \ (V_1 \subseteq V_2)$.
- $AV_1 =_a AV_2$, iff $(AV_1 \leq_a AV_2) \wedge (AV_2 \leq_a AV_1)$.
- $AV_1 <_a AV_2$ (or $AV_2 >_a AV_1$), iff $(AV_1 \leq_a AV_2) \wedge (AV_1 \neq_a AV_2)$.

According to Definition 1, for any instantiated context AV, $AV \leq_a *$.

The $=_a$ relationship between two instantiated contexts AV_1 and AV_2 indicates that they both have exactly the same contextual attributes with the same attribute values. $AV_1 \leq_a AV_2$ states that AV_2 is broader than AV_1, covering the contextual scope of AV_1.

Example 2. Assume we have four instantiated contexts:
$AV_1 = *$, $AV_2 = \{Year:=\{“1999”, “2000”, “2001”\}\}$,
$AV_3 = \{Region:=\{“north”\},Year:=\{“2000”\}\}$,and
$AV_4 = \{Region:=\{“south”\}, Year:=\{“2000”\}\}$.
Following Definition 2, we have $AV_2 <_a AV_1$, $AV_3 <_a AV_1$, $AV_4 <_a AV_1$, $AV_3 <_a AV_2$, and $AV_4 <_a AV_2$.

Based on the primitive relationships of concepts (*Is-A*, *Synonym*, *Opposite*), as well as their contexts ($\leq_a$,$=_a$,$<_a$), we can formulate the following three concept-term-based relationships, i.e., equivalence EQ_{ct}, specification $SPEC_{ct}$ and opposition $OPSI_{ct}$. Assume $n_1=e_1|_{AV1}$ and $n_2= e_2|_{AV2}$ are two concept terms in the following definitions, where $n_1, n_2 \in CTerm$.

Definition 3. **Equivalence** EQ_{ct} (n_1, n_2). n_1 is **equivalent** to n_2, iff the following two conditions hold: 1) $(e_1= e_2) \vee$ Synonym (e_1, e_2); 2) $(AV_1=_a AV_2)$.

Example 3. Given two concept terms: n_1 = *wet-winter*$|_{\{Region:=\{“north”\}\}}$ and n_2=*high-rainfall-winter*$|_{\{Region:=\{“north”\}\}}$, EQ_{ct} (n_1, n_2) since Synonym (*wet-winter*, *high-rainfall-winter*).

Definition 4. **Specification** $SPEC_{ct}$ (n_1, n_2). n_1 is a **specification** of n_2 (conversely, n_2 is a **generalization** of n_1), iff the following two conditions hold:
1) $(e_1= e_2) \vee$ Is-A $(e_1, e_2) \vee$ Synonym (e_1, e_2); 2) $(AV_1 \leq_a AV_2)$.

Example 4. Let n_1 = $wet\text{-}winter|_{\{Region:=\{“north”\},\ Year:=\{“2000”\}\}}$,
n_2= $wet\text{-}winter|_{\{Year:=\{“2000”\}\}}$ be two concept terms. $SPEC_{ct}$ (n_1, n_2) since $\{Region:=\{“north”\},\ Year:=\{“2000”\}\} <_a \{Year:=\{“2000”\}\}$.

Definition 5. **Opposition** $OPSI_{ct}$ (n_1, n_2). n_1 is **opposite** to n_2, iff the following two conditions hold 1) Opposite (e_1, e_2); 2) ($AV_1 =_a AV_2$).

Example 5. Let n_1= $wet\text{-}winter|_{\{Region:=\{“north”\}\}}$, n_2= $dry\text{-}winter|_{\{Region:=\{“north”\}\}}$ be two concept terms. $OPSI_{ct}$ (n_1, n_2) since Opposite(*wet-winter, dry-winter*).

To facilitate the description of hypothesis-based inter-relationships in Subsection 3.4, we further extend the three relationships (i.e., EQ_{ct}, $SPEC_{ct}$ and $OPSI_{ct}$) defined over a pair of concept terms to the ones (i.e., EQ_{ct}, $SPEC_{ct}$ and $OPSI_{ct}$) over a pair of concept term sets. Let N_1 and N_2 be two concept term sets.

Definition 6. **Equivalence** EQ_{CT} (N_1, N_2). N_1 is **equivalent** to N_2, iff $\forall n_1 \in N_1\ \exists n_2 \in N_2\ EQ_{ct}(n_1, n_2) \wedge \forall n_2 \in N_2\ \exists n_1 \in N_1\ EQ_{ct}(n_2, n_1)$.

Definition 7. **Specification** $SPEC_{CT}$ (N_1,N_2). N_1 is a **specification** of N_2, iff $\forall n_2 \in N_2\ \exists n_1 \in N_1\ SPEC_{ct}(n_2, n_1)$.

Definition 8. **Opposition** $OPSI_{CT}$ (N_1,N_2). N_1 is **opposite** to N_2, iff $\exists n_1 \in N_1\ \exists n_2 \in N_2\ OPSI_{ct}(n_1,n_2) \vee \exists n_2 \in N_2\ \exists n_1 \in N_1\ OPSI_{ct}(n_2,n_1)$.

As long as there exists a pair of opposite concept terms in the two concept term sets, we declare they are opposite.

3.3 Relation Terms

A relation concept explicates a certain correlation among a set of conceptual terms. Unlike entity concepts, relation concepts can be affiliated with different kinds of modals like *necessity, possibility, permission*, etc. to qualify the truth of the relationships. In the paper, we apply well-established modal logic [4] to our relation concept study. By prefixing a relation concept *r* with the symbol   or ◊, we can achieve different levels of ascertain ability regarding relation *r*. For example,   *cause* implies a *necessarily causal* relation, while ◊ *cause* implies a *possibly causal* relation.

Definition 9. A **relation term** *m* is of the form $m = \delta r$, where $r \in RConcept$, and δ could be , ◊, or an empty modal.

Definition 10. According to modal logic, we define the order “ ” < “ ” < “◊” for symbols , empty modal, and ◊.

The three relation-term-based relationships can be defined using the same names (i.e., *EQ, SPEC* and *OPSI*) as concept-term-based relationships but with a different subscript flag "$_{rt}$" to make the difference.

Definition 11. **Equivalence** EQ_{rt} $(\delta_1 r_1,\ \delta_2 r_2)$. $\delta_1 r_1$ is **equivalent** to $\delta_2 r_2$, iff the following two conditions hold: 1) $(r_1 = r_2) \vee$ Synonym (r_1, r_2); 2) $(\delta_1 = \delta_2)$.

Definition 12. **Specification** $SPEC_{rt}$ $(\delta_1 r_1,\ \delta_2 r_2)$. $\delta_1 r_1$ is a **specification** of $\delta_2 r_2$ (conversely, $\delta_2 r_2$ is a **generalization** of $\delta_1 r_1$), iff the following two conditions hold:
1) $(r_1 = r_2) \vee$ Is-A $(r_1, r_2) \vee$ Synonym (r_1, r_2); 2) $(\delta_1 = \delta_2) \vee (\delta_1 < \delta_2)$.

Definition 13. Opposition $OPSI_{rt}$ $(\delta_1 r_1,\ \delta_2 r_2)$. $\delta_1 r_1$ is **opposite** to $\delta_2 r_2$, iff the following two conditions hold: 1) Opposite (r_1, r_2); 2) $(\delta_1 = \delta_2 \neq$ "◊").

Example 6. EQ_{rt} (◊*cause*, ◊*lead-to*), $SPEC_{rt}$ (*cause*, ◊*cause*), $SPEC_{rt}$ (*cause*, *relate*), and $OPSI_{rt}$ (*relate*, *unrelate*).

3.4 Hypotheses

A hypothesis communicates a human's cognitive idea or thinking about things in existence, such as the causal connection of situations, the sequential occurrence of events, etc. Here, we describe each piece of hypothesis using a relation concept which correlates a set of input concept terms to a set of output concept terms. For example, the hypothesis "*wet winter in the north causes summer flood in the south and hot summer in the east*" causally relates concept term $\textit{wet-winter}|_{\{Region:=\{\text{"north"}\}\}}$ to $\textit{summer-flood}|_{\{Region:=\{\text{"south"}\}\}}$ and $\textit{hot-summer}|_{\{Region:=\{\text{"east"}\}\}}$.

Definition 14. A hypothesis H is of the form $H=\delta r(I_N, O_N)$, where δr is a relation term, I_N and O_N are concept term sets.

Various hypothesis-based inter-relationships can be established based on the relationships of their components, i.e., concept term sets and relation terms. Assume that $H_1=\delta_1 r_1\ (I_{N1}, O_{N1})$ and $H_2=\delta_2 r_2\ (I_{N2}, O_{N2})$ are two hypotheses.

Definition 15. H_1 is **equivalent** to H_2, written as $H_1 =_{\mathrm{h}} H_2$, iff $EQ_{rt}\ (\delta_1 r_1,\ \delta_2 r_2) \wedge EQ_{CT}\ (I_{N1}, I_{N2}) \wedge EQ_{CT}\ (O_{N1}, O_{N2})$.

For two equivalent hypotheses, they must have equivalent relation terms, as well as equivalent input and output concept term sets.

Definition 16. H_1 is a **specification** of H_2 (conversely, H_2 is a **generalization** of H_1), written as $H_1 \leq_{\mathrm{h}} H_2$ $(H_2 \geq_{\mathrm{h}} H_1)$, iff $SPEC_{rt}\ (\delta_1 r_1,\ \delta_2 r_2) \wedge SPEC_{CT}\ (I_{N1}, I_{N2}) \wedge SPEC_{CT}\ (O_{N1}, O_{N2})$.
We call H_1 a **strict specification** of H_2 (conversely, H_2 a **strict generalization** of H_1), written as $H_1 <_{\mathrm{h}} H_2$ $(H_2 >_{\mathrm{h}} H_1)$, iff $(H_1 \leq_{\mathrm{h}} H_2) \wedge (H_1 \neq_{\mathrm{h}} H_2)$.

If H_1 is a specification of H_2 and a specification of H_3, then H_1 is a **common specification** of H_2 and H_3. Conversely, if H_1 is a generalization of H_2 and a generalization of H_3, then H_1 is a **common generalization** of H_2 and H_3.

Example 7. Given the following three hypotheses:
H_1 = *cause* ({*wet-winter*$|_{\{Region:=\{\text{“north”}\}\}}$, *warm-winter*$|_{\{Region:=\{\text{“north”}\}\}}$}, {*summer-flood*$|_{\{Region:=\{\text{“south”}\}\}}$}),
H_2 = ◊*cause* ({*wet-winter*$|_{\{Region:=\{\text{“north”}\}\}}$ }, {*summer-flood*$|_{\{Region:=\{\text{“south”}\}\}}$}),
H_3 = ◊*cause* ({*wet-winter*$|_{\{Region:=\{\text{“north”}\}\}}$ }, {*river-behavior*$|_*$}),
H_1 is more specific than H_2 and H_3, and H_2 is also more specific than H_3 (i.e., $H_1 \leq_h H_2$, $H_1 \leq_h H_3$ and $H_2 \leq_h H_3$). All of them are strict specifications. Besides, H_1 is a common specification of H_2 and H_3, and H_3 is a common generalization of H_1 and H_2.

Definition 17. H_1 is **opposite to** H_2, written as $H_1 \propto_h H_2$, iff either of the following conditions holds:

1) $OPSI_{rt}\ (\delta_1 r_1,\ \delta_2 r_2) \wedge EQ_{CT}\ (I_{N1},\ I_{N2}) \wedge EQ_{CT}\ (O_{N1},\ O_{N2})$; or
1) $EQ_{rt}\ (\delta_1 r_1,\ \delta_2 r_2) \wedge EQ_{CT}\ (I_{N1},\ I_{N2}) \wedge OPSI_{CT}\ (O_{N1},\ O_{N2})$.

For two opposite hypotheses, they may have equivalent input/output concept term sets but with opposite relation terms (Case 1 of the definition), or they may have equivalent relation terms and input concept term sets, but with at least one opposite output concept term pair (Case 2 of the definition).

Example 8. Given the following two hypotheses:
H_1 = *relate*({*wet-winter*$|_*$}, {*summer-flood*$|_*$, *hot-summer*$|_*$}),
H_2 = *unrelate*({*wet-winter*$|_*$}, {*summer-flood*$|_*$, *hot-summer*$|_*$}),
$H_1 \propto_h H_2$ since $OPSI_{rt}$ (*relate*, *unrelate*).

3.5 The Knowledge Subspace and its Linkage to the Document Subspace

Hypotheses and their inter-relationships constitute a DL knowledge subspace. At an abstract level, a knowledge subspace can be viewed as an oriented diagram consisting of a series of nodes (each representing a hypothesis) that are connected to each other through directed labeled edges (representing various relationships between hypotheses), as shown in the upper part of Figure 2. To make the diagram connected, we introduce two special hypotheses: the universal hypothesis ~ that is a generalization of all other hypotheses, and the absurd hypothesis ⊥ that is a specification of all other hypotheses.

Definition 18. A DL **knowledge subspace** is composed of a set of nodes representing hypotheses, and a set of directed edges representing relationships of hypotheses.

The DL document subspace accommodates all the documents in a library. They are the sources for answering users' information searching and browsing requests. In addition, for an enhanced DL system proposed in this paper, documents in the DL document subspace also serve as the justification for the corresponding knowledge in the knowledge subspace. That is, under each hypothesis is a set of justification documents, giving reasons and evidences for that knowledge. Let *Doc* denote the whole set of documents in a DL. All the documents in the DL that support a hypothesis constitute the referent for that hypothesis.

Definition 19. Let H be a hypothesis in the knowledge subspace. The **referent** of H, written as φH, is the set of documents in the library that support H. For the two special hypotheses, we assume that $\varphi\perp = \emptyset$ and $\varphi\sim = Doc$.

The defined equivalence, (strict) specification and opposition relationships of hypotheses lead us to the following axiom and theorem.

Axiom 1. Let H_1, H_2 be two hypotheses.

– If H_1 a (strict) specification of H_2, i.e., $H_1 \leq_h (<_h) H_2$, then $\varphi H_1 \subseteq \varphi H_2$.
– If H_1 equivalent to H_2, i.e., $H_1 =_h H_2$, then $\varphi H_1 = \varphi H_2$.
– If H_1 opposite to H_2, i.e., $H_1 \propto_h H_2$, then $\varphi H_1 \cap \varphi H_2 = \emptyset$.

Using Definition 16 of specialization/generality between hypotheses, we can be sure that if a hypothesis is consistent with a set of documents, any generalization of it will also be consistent with this document set. In contrast, if a document does not conform to a hypothesis, it cannot conform to any specialization of that hypothesis either. For any hypothesis $H \in Hypo$ where $(\perp \leq_h \mathrm{H} \leq_h \sim)$, it is obvious that

$\emptyset = \varphi\perp \subseteq \varphi H \subseteq \varphi\sim = Doc$.

Theorem 1. Let H_1, H_2, H_3 be three hypotheses, where H_1 is a common generalization of H_2 and H_3, i.e., $(H_2 \leq_h H_1)$ and $(H_3 \leq_h H_1)$. We have $(\varphi H_2 \cup \varphi H_3) \subseteq \varphi H_1$.

Proof. Since $(H_2 \leq_h H_1)$, according to Axiom 1, $(\varphi H_2 \subseteq \varphi H_1)$. Similarly, $(\varphi H_3 \subseteq \varphi H_1)$ because of $(H_3 \leq_h H_1)$. Thus, $(\varphi H_2 \cup \varphi H_3) \subseteq \varphi H_1$.

4. Conclusion

In this paper, we present a two-layered DL model to address human's different information requirements. On the basis of the proposed model, we divide the DL information space into a knowledge subspace and a document subspace. A detailed description of the knowledge subspace and its construction mechanisms, as well as query facilities against the enhanced DLs are particularly discussed. Currently, we are researching practical methods of knowledge acquisition to fill in the knowledge subspace.

References

1. M.J.Bates. Subject access in online catalogs: A design model. Journal of the American Society for Information Science, 37(6):357-376, 1986.
2. K.Beard and V.Sharman. Multidimensional ranking in digital spatial libraries. Proc. of the 2nd Intl. Conf. On the Theory and Practice of Digital Libraries, June 1995.
3. P.Buneman, A.Deutsch, and W.C.Tan. A deterministic model for semi-structured data. Proc. of the 1999 Intl. Workshop on Query Processing for Semi-Structured Data and Non-Standard Data Formats, January 1999.
4. B.F.Chellas. Modal Logic:An Introduction. Cambridge University Press, 1980.
5. H.Chen, K.J.Lynch, K.Basu, and T.Ng. Generating, integrating, and activating thesauri for concept-based document retrieval. IEEE Expert, 8(2):25-34, 1993.
6. M.Dunlop and C. van Rijsbergen. Hypermedia and free text retrieval. Journal of Information Processing and Management, 29(3), 1993.
7. C.E.Dyreson, M.H.Bohlen, and C.S.Jensen. Capturing and querying multiple aspects of semistructured data. Proc. of the 1999 Intl. Conf. VLDB, pages 290-301, 1999.
8. M.W.Evens and R.N.Smith. A lexicon for a computer question-answering system. American Journal of Computational Linguistics, 83:1-93,1979.
9. W.B.Frakes and R.Baeza-Yates. Information Retrieval: Data Structures and Algorithms, Prentice Hall, 1992.
10. T.Gruber. A translation approach to portable ontologies. Knowledge Acquisition, 5(2):199-220, 1993.
11. N.Guarino and C.Welty. Ontological analysis of taxonomic relationships. Proc. of the 19th Intl. Conf. Conceptual Modeling, 2000.
12. F.Kiefer editor. Studies in Syntax and Semantics, chapter Semantics and Lexicography: Towards a New Type of Anilingual Dictionary, 1969.
13. S.Melnik, H.Garcia-Molina, and A.Paepcke. A mediation infrastructure for digital library services. Technical report, Stanford University, 2000.
14. J.T.Nutter, E.A.Fox, and M.W.Evens. Building a lexicon from machine-readable dictionaries for improved information retrieval. Journal of Literary and Linguistic Computing, 5(2):129-137,1990.
15. A.Paepcke, R.Brandriff, G.Janee, R.Larson, B.Ludaescher, S.Melnik, and S.Raghavan. Search middleware and the simple digital library interoperability protocol. D-Lib Magazine, 6(3), 2000.
16. A.Paepcke, s.b.cousins, H.Garcia-Molina, S.W.Hassan, s.k.Ketchpel, M.Roscheisen, and T.Winograd. Using distributed objects for digital library interoperability. Technical report, Stanford University, 1998.
17. M.Papazoglou and J.Hoppenbrouwers. Contextualizing the information space in federated digital libraries. SIGMOD Record, 28(1):40-46, 1999.
18. S.Payette and C.Lagoze. Flexible and extensible digital object and repository architecture (FEDORA). Proc. of the 2nd European Conf. On Research and Advanced Technology for Digital Libraries, pages 41-59, 1999.
19. J.F.Sowa. Conceptual Structures: Information Processing in Mind and Machine. Addison-Wesley, 1984.
20. B.Velez, R.Weiss, M.A. Sheldon, D.K.Gifford. Fast and effective query refinement. Proc. of the 20th Intl. ACM SIGIR Conf. On Research and Development in Information Retrieval, pages 6-15, 1997.
21. P.Willett. Recent trends in hierarchical document clustering: a critical review. Information Processing and management, 24(5), 1998.